① Expl = X-var ← set/determined by researcher
Resp = Y-var ← measured

Ex: mg. of a drug & # of tumors developed

② 2 quant. var's (#)

categorical var - color/shape

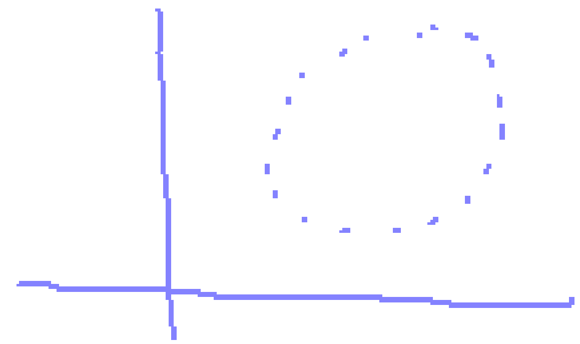
③ r = correlation coefficient

decimal

measures strength of linear relationship

$-1 \le r \le 1$

$r = 0$ = no linear relat.



$r \approx -1, 1$ = strong linear

r^2 = coefficient of determination

%

___% of the change in y-var. that was due to the change in the x-var.

4

5 Form: linear, curved, scattered

direction: + or — or scattered

strength: strong
moderate
weak
Scattered

6. LSR =

$$\hat{y} = a + bx$$

LSR = least squares regression line

* avg. of scatterplot

$(\bar{x}, \bar{y})$

8: LinReg(a+bx) X, Y, $\underline{Y1}$ ← puts line in y= ← prediction

Y1(100)

a =

b =

r =

r^2 =

⑦ Residuals = errors btw. LSR line + actual datapt.

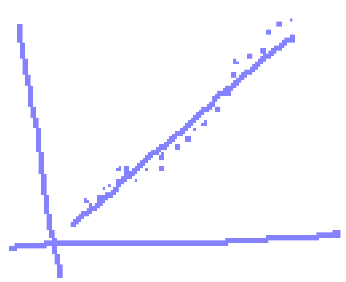
= obs. − exp

actual data pt.   LSR line

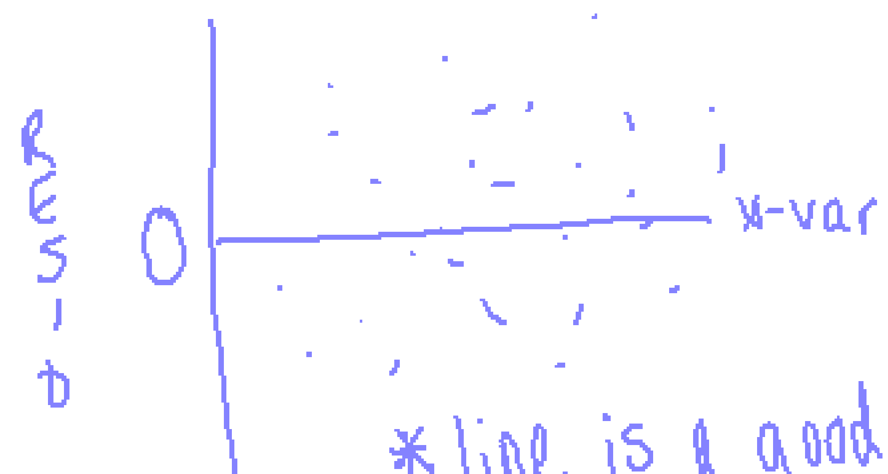
Residual Plot:

x-var. vs. RESID

* y-direction



line is not the best fit for data

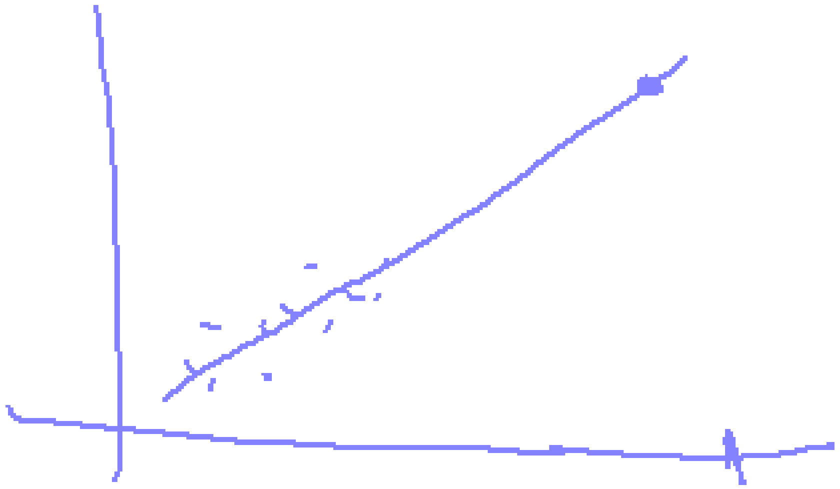


* line is a good fit for data

wksht - Airfares

8. For every 1 mile traveled, the airfare increased on avg. by \$0.12

11. extrapolation



10.1 - testing the slope

sample

LSR line

$\hat{y} = a + bx$

$\hat{p}$

$a = \bar{y} - b\bar{x}$

$b = r\frac{S_y}{S_x}$

Book/Form sheet

$\hat{y} = b_0 + b_1 x$

$b_0 = \bar{y} - b_1\bar{x}$

$b_1 = r\frac{S_y}{S_x}$

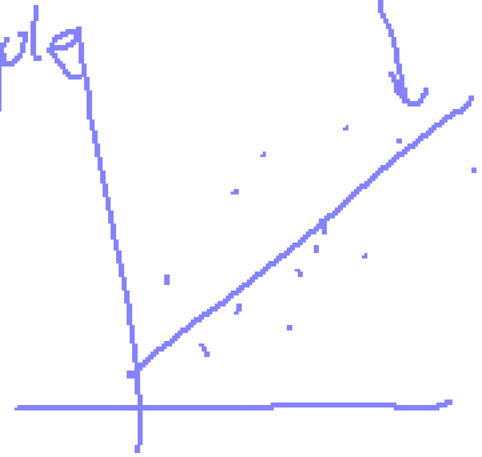
stats sample	pop. param
$\hat{y}$	y
a	α
b	β ← pop. slope
e_i errors	ε_i

β_0

β_1

Sample

LSR



pop. line of best fit

sample

LSR line basis for inference about a population where our data is a sample.

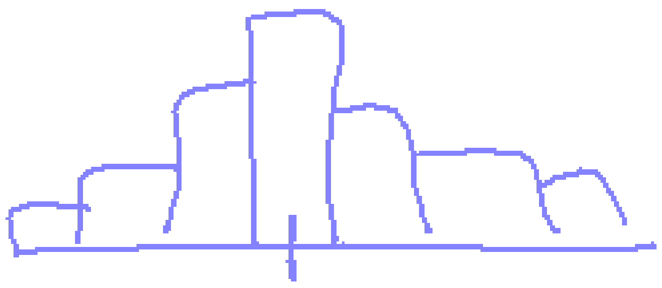
LSR line: $\hat{y} = a + bx + e_i$

pop. line: $y = \alpha + \beta x + \varepsilon_i$

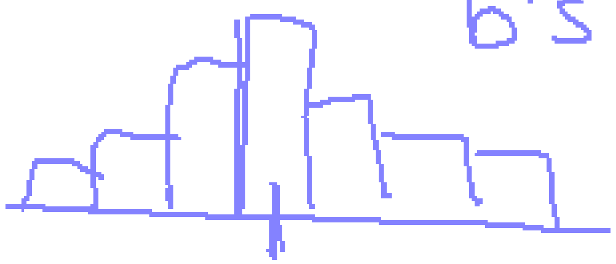
a and b

- unbiased estimators of α and β
- normally distributed
  centers @ α and β

a's



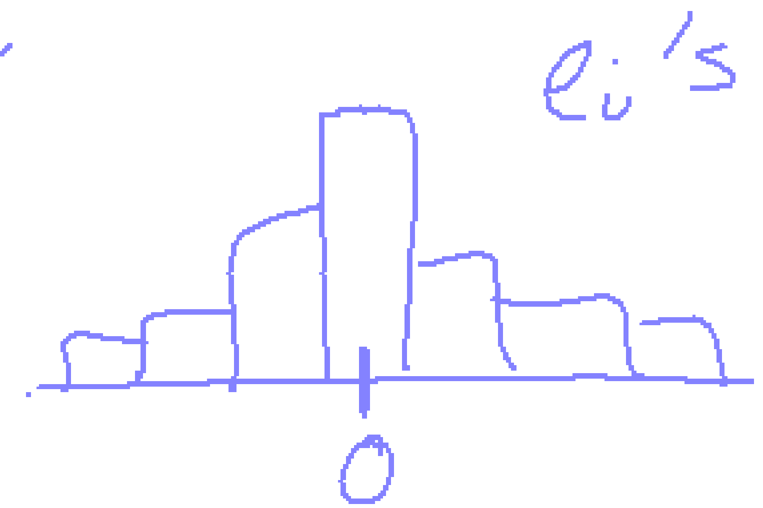
b's



ε_i = residuals = deviations

- estimated by e_i
- independent
- normally distrib.
- mean = 0

  std. dev. of residuals = σ
- resid = obs. − predicted
- $\sum e_i = 0$



pop.

σ = std. dev. of resid.

- estimated by s

resid

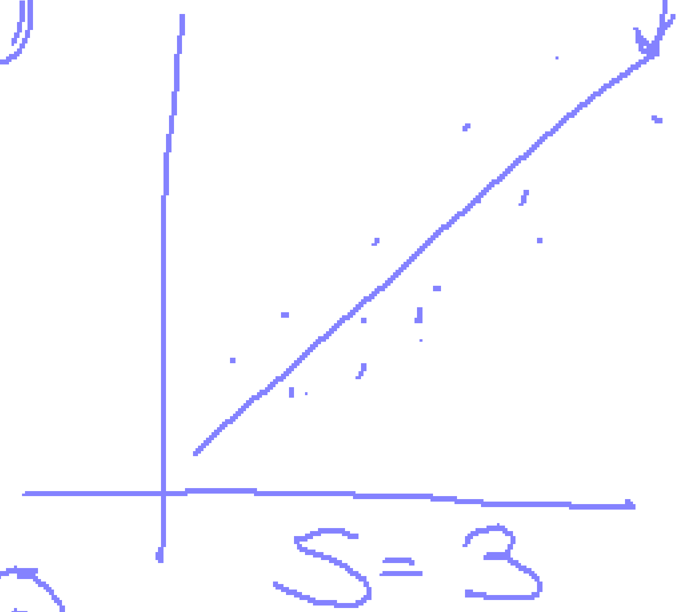
- $s = \sqrt{\dfrac{\sum e_i^2}{n-2}}$ ← 2 vars.

- $df = n-2$

estimate of $\alpha = a$

$\beta = b$

① center/mean



②